

PHIL 7001: Fundamentals of AI, Data, and Algorithms

Lecture 1

Risk and Uncertainty: Subways, Coconuts and Black Swans

Boris Babic,

HKU 100 Associate Professor of Data Science, Law and Philosophy



Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

Introduction

About the Class

Boris
Babic,
HKU

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

- An introduction to the core quantitative concepts behind AIES.
- Goal: how to intelligently think about and use data driven systems.
- Please regularly consult the syllabus on Moodle. I will update it if necessary.

About Us

Boris
Babic,
HKU

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

- Boris Babic is associate professor of data science, law and philosophy at HKU.
- His research is primarily in law and ethics of machine learning and AI.
- Maomei Wang is a PhD student and instructor in the Faculty of Arts.
- Her research is primarily in decision theory and formal epistemology.

Boris
Babic,
HKU

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

Uncertainty

- Much of Statistics, machine learning, and AI is about forecasting/prediction.
- ML is about classification (a type of prediction).
- NLP is oriented toward language, but it is in itself a type of prediction.
- And statistics is at bottom about inference and prediction.
- Hence, at the core of everything we will do is statistical inference and prediction. This will be our starting point.
- Inference and prediction are only important under conditions of uncertainty. If we knew the future deterministically, we wouldn't care!
- Today's class will be a big picture introduction for how to think about statistics, inference, and modeling.
- In the classes to follow, we will get into the quantitative details.

Forecasting, Prediction, Classification

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise



SEARCH

FORTUNE Well.

Home Life Health Mind Family Aging Well

HEALTH · ECONOMY

The COVID-19 pandemic cost the U.S. economy \$14 trillion, new research finds

BY JAKUB HLÁVKA, ADAM ROSE AND THE CONVERSATION

May 17, 2023 at 12:24 AM GMT+8



Forecasting, Prediction, Classification

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

The Social and Political Costs of the Financial Crisis, 10 Years Later

by Gautam Mukunda

September 25, 2018



NurPhoto/Getty Images

Summary. The economic costs of the financial crisis were staggering. But the most important effects of the financial crisis may be political and social, not economic. The years after the crisis saw sharp increases in political polarization and the rise of populist movements on both... [more](#)

It is hard to overstate the sheer economic cost of the 2008 financial crisis. The combination of increased expenditures and decreased revenues resulting from the crisis from 2008 to 2010 is likely to cost the United States government well over \$2 trillion, more than twice the cost of the 17-year-long war in Afghanistan. Broader measures are

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

US general's 'gut' feeling of war with China sparks alarm over predictions

Leaked memo forecasting Taiwan strait conflict in 2025 triggers debate about 'undisciplined' comments



Forecasting, Prediction, Classification

Boris
Babic,
HKU

Introduction

Uncertainty

Examples

Rare events

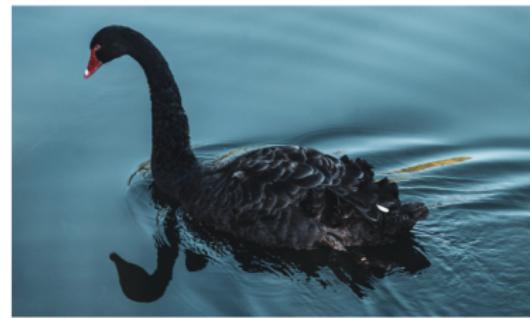
25 Sigma

Models

The Illusion
of Control

Risk

Exercise



- Nasim Taleb: black swans are extremely rare events of disproportionate consequence.
- A problem for standard statistical modeling?
- Are these black swans?

Forecasting, Prediction, Classification

Introduction

Uncertainty

Examples

Rare events

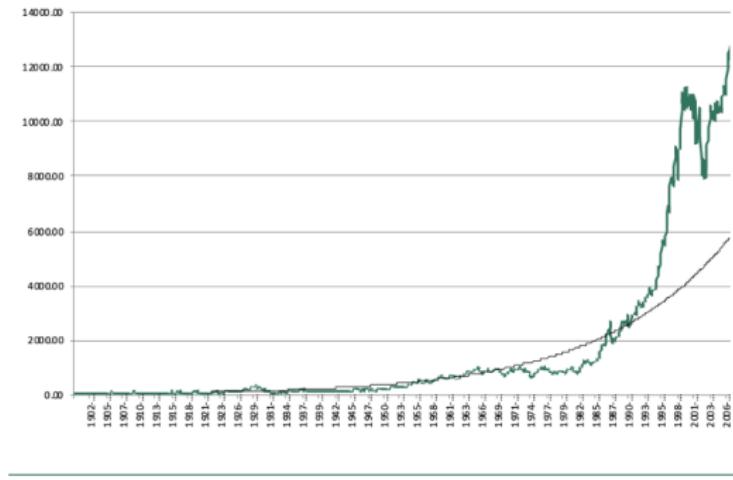
25 Sigma

Models

The Illusion
of Control

Risk

Exercise

THE 2007 BUBBLE DJIA

Forecasting, Prediction, Classification

Introduction

Uncertainty

Examples

Rare events

25 Sigma

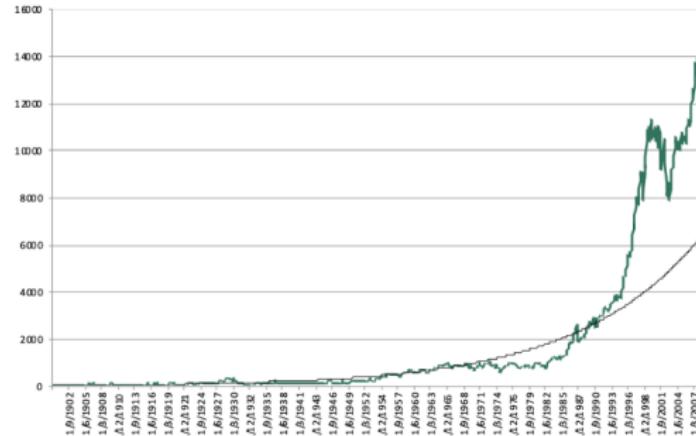
Models

The Illusion
of Control

Risk

Exercise

2008-09 : BURSTING OF THE 2007 BUBBLE DJIA



Forecasting, Prediction, Classification

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

THE 1966 BUBBLE DJIA

Forecasting, Prediction, Classification

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

EARLY 1980s : THE BURSTING OF THE 1966 BUBBLE DJIA



Forecasting, Prediction, Classification

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

THE 1929 BUBBLE DJIA

Forecasting, Prediction, Classification

Boris
Babic,
HKU

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

THE 1930S: THE BURSTING OF THE 1929 BUBBLE DJIA



Forecasting, Prediction, Classification

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

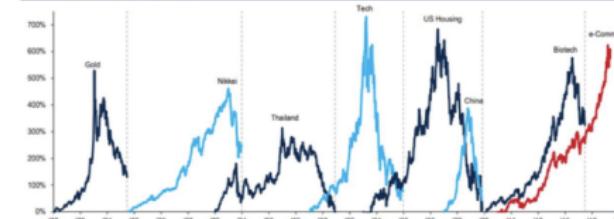
Not Black Swans!

Equity Bubbles



- Dow Jones eCom index (Amazon, Netflix, Google and Facebook) up 617%, 3rd largest bubble of past 40 years
- U.S. tech market cap (\$6.0tn) exceeding that of all companies in the Eurozone (\$5.0 tn)
- Facebook (25k employees) market cap > MSCI India (1.3 bn people)

Chart 8: Asset price bubbles of the past 40 years



Source: BofA Merrill Lynch Global Investment Strategy, Bloomberg, New Gold 2002 (Currencies), Japanese Equities Nikkei Index, The S&P 500 (S&P 500 Index), US Housing (DOWJONES Index), Commodity (DCCMP), Velocity Biotech (BMO Financial Group), EQUITYCMX Index.

Gold = Gold spot price in U.S. dollars, Nikkei = Nikkei 225 stock average in a price weighted index of 225 stocks listed on the Tokyo Stock Exchange, Thailand = Thailand Composite Stock Price Index, Tech = NASDAQ 100 stock index = cap-weighted index of 100 largest stocks on the U.S. NASDAQ exchange, US Housing = S&P Case-Shiller U.S. National Home price index tracks the value of single-family homes in the United States, China = Shanghai Composite Index, Biotech = Nasdaq Biotechnology Index, e-commerce = Dow Jones Internet Commerce Index (a modified cap-weighted index that tracks the performance of companies involved in internet commerce, you cannot invest directly in an index).

0-02-18-TR-Weston

Forecasting, Prediction, Classification

Boris
Babic,
HKU

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

- 2019: Covid 19
- 2012: MERS
- 2009: H1N1
- 2002: SARS
- 1968: H3N2
- 1957: H2N2
- 1918: H1N1 (50 million deaths/ 500 million cases)
- Not a black swan!

The Reference Class Problem

Boris
Babic,
HKU

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

- At what level of generality do we frame the historical reference class?
- If we ask how many human health outbreaks caused by SARS-CoV-2 – the unique strain of coronavirus causing Covid 19 – then Covid 19 is singular.
- If we ask, how many human health outbreaks caused by a coronavirus, then it is one of a handful.
- If we ask, how many human health outbreaks caused by a virus, then it is one of many.
- Which of these is right?
- This is known as **the reference class problem**.
- You cannot *do the probability math* without first using your judgment and making a call about this!

How Unlucky is 25 Sigma?



Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

- In August 2007, then CFO of Goldman Sachs, David Viniar, announced that Goldman's flagship GEO hedge fund had lost 27% of its value since the start of the year.
- As Mr. Viniar explained: We were seeing things that were 25-standard deviation moves, several days in a row.
- Question 1: What is the probability of a financial loss that is 2 standard deviations away from the mean within a normal model? (i.e., a 2 Sigma event)?
- Question 2: Can you think of an event – any event! – whose probability of occurrence is 20-25 standard deviations away from the mean?

How Unlucky is 25 Sigma?

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

Table 1: Probabilities of k -sigma events: $k=3, 4, 5, 6$ and 7

k	Probability in any given day	Expected occurrence: once in every	
3	0.135%	740.8	days
4	0.00317%	31559.6	days
5	0.000029%	3,483,046.3	days
6	0.000000099%	1,009,976,678	days
7	0.000000000129%	7.76e+11	days

How Unlucky is 25 Sigma?

[Introduction](#)[Uncertainty](#)[Examples](#)[Rare events](#)[25 Sigma](#)[Models](#)[The Illusion
of Control](#)[Risk](#)[Exercise](#)

- A 5-sigma event corresponds to an expected occurrence of less than just one day in the entire period since the end of the last Ice Age.
- A 6-sigma event corresponds to an expected occurrence of less than one day in the entire period since our species, Homo Sapiens, evolved from earlier primates.
- An 8-sigma event is equal to an expected occurrence of just once in a period which is considerably longer than the time elapsed since Big Bang.
- A 20-sigma event is equal to an expected occurrence of one day in a number of years equal to 10 times the likely number of particles in the universe.

Modeling Choices

Boris
Babic,
HKU

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

- ***Complex Models*** can fit past data well but don't necessarily predict well.
- ***Simple Models*** don't necessarily fit past data well but predict the future better than ***Complex Models***.
- ***Combining predictions across models*** improves accuracy
- However, in socio-economic domains, despite the current *best approach*, our prediction power remains limited...

The Most Famous Aphorism in All of Statistics

Boris
Babic,
HKU

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

*All models are wrong
but some are useful*



George E.P. Box

The Problem of Individual Heterogeneity in Human Sciences

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

PHYSICS ENVY IN SOCIO-ECONOMIC MODELS:

The more mathematically sophisticated the model



The better we can predict, the better we can control risk and uncertainty,...

Economists suffer from a deep psychological disorder that I call 'physics envy'. We wish that 99 percent of economic behavior could be captured by three simple laws of nature. In fact, economists have 99 laws that capture 3 percent of behavior. Economics is a uniquely human endeavor.

Andrew Lo

MIT Professor of financial economics

Economists who adhere to rational-expectations models of the world will never admit it, but a lot of what happens in markets is driven by pure stupidity - or, rather, inattention, misinformation about fundamentals, and an exaggerated focus on currently circulating stories.

Robert Shiller
Nobel Laureate in 2013

Wall Street indices predicted nine out of the last five recessions !

Paul Samuelson
Nobel Laureate in 1970

Ensembles: Combining Models

Introduction

Uncertainty

Examples

Rare events

25 Sigma

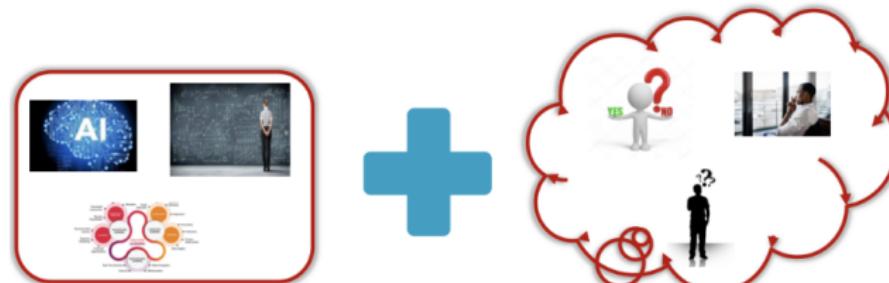
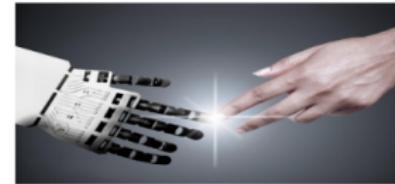
Models

The Illusion
of Control

Risk

Exercise

**COMBINING JUDGMENTS ACROSS MODELS IMPROVES ACCURACY:
MACHINE + HUMAN COMBINATION**



Probability vs. Statistics?

Introduction

Uncertainty

Examples

Rare events

25 Sigma

Models

The Illusion
of Control

Risk

Exercise

- What is the difference between a probability question and a statistics question?
- I will toss a fair coin twice. What is the likelihood of observing two heads? What kind of question is this?
- I tossed a certain coin 3 times, and observed the results: heads, heads, tails. What is the likelihood that the coin is fair? What kind of question is this?
- Probability is analytic and deductive; statistics is synthetic and inductive.
- Probability is objective, statistics is not. Statistical procedures are objective, but there is no “mathematically” unique answer that determines which model is right, which procedure is appropriate, etc.

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

The Illusion of Control

Low Hanging Fruit and The Illusion of Control

Introduction

Uncertainty

**The Illusion
of Control**

Risk

Exercise



- The biggest problem in machine learning and AI: too easy to focus on the low hanging fruit!

The Illusion of Control: Life Expectancy and Childhood Mortality

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

Once the dangerous childhood years were passed, however, Victorian contemporary sources reveal that life expectancy in the mid-Victorian period was not markedly different from what it is today. Once infant mortality is stripped out, life expectancy at age five was 75 for men and 73 for women.

– J Rowbotham and P Clayton (2008), *J R Soc. Med.* 101(9): 454–462.

The Illusion of Control

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

NYT, January 8, 2015 – By Ezekiel J. Emanuel

The Opinion Pages | CONTRIBUTING OP-ED WRITER

Skip Your Annual Physical

JAN. 8, 2015



Tom Toles

WE all make resolutions and promises to live healthier and better lives, to make the world a better place. Not having my annual physical is one small way I can help reduce health care costs — and save myself time, worry and a worthless exam.

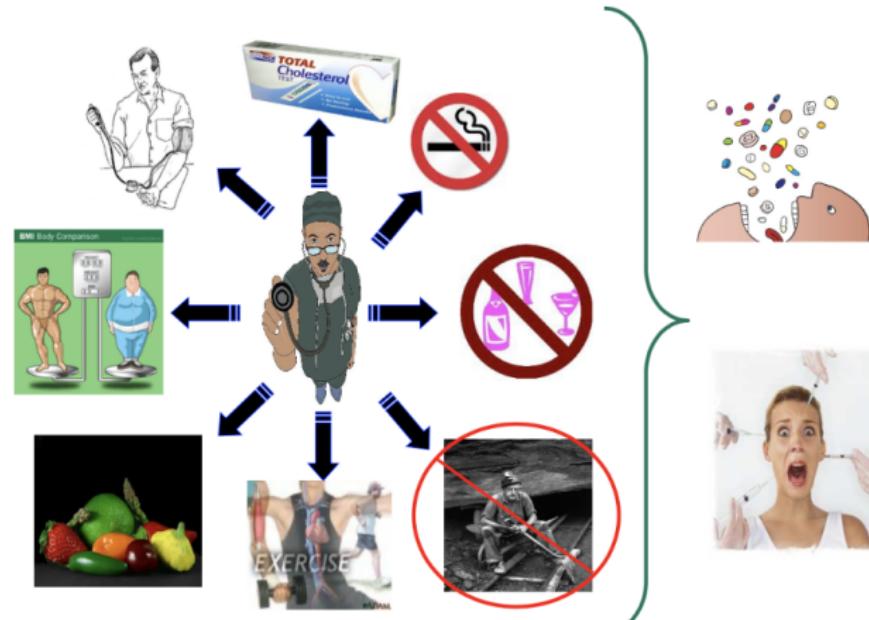
Around 45 million Americans are likely to have a routine physical this year — just as they have for many years running. A poke here, a listen there, a few tubes of blood, maybe an X-ray, a few reassuring words about diet, exercise and not smoking from the doctor, all just to be sure everything is in good working order. Most think of it as the human equivalent of a 15,000-mile checkup and fluid change, which can uncover hidden problems and ensure longer engine life.

There is only one problem: From a health perspective, the annual physical exam is basically worthless.

Ezekiel J. Emanuel, an oncologist and former White House adviser, is a vice provost and professor at the University of Pennsylvania. He is a contributing opinion writer for The New York Times on a range of topics including health and health policy.

There is only one problem: From a health perspective, the annual physical exam is basically worthless.

The Illusion of Control



The Illusion of Control

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

PREDICTIONS: MEDICINE

Years of life expectancy gained

Meta Analysis

Major risk factor	Contribution	(Lancet study)	(Harvard study)
1. High blood pressure	4.4%	0.90	0.75
2. High cholesterol	2.8%	0.57	0.75
3. High BMI (Body Mass Index)	2.3%	0.47	0.60
4. Low fruit/vegetable intake	1.8%	0.37	0.37*
5. Physical inactivity	1.3%	0.26	0.62
6. Tobacco	4.1%	0.84	0.75
7. Alcohol	4.0%	0.81	0.81*
8. Occupational risk factors	0.9%	0.18	0.18*
Total	21.6%	4.40 years	4.83 years

Consequences of The Illusion of Control

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

- ACE inhibitors – kidney impairment
- Alpha blockers – Orthostatic hypertension, depression
- Beta blockers – sexual dysfunction
- Calcium channel blockers – toxicity, peripheral edema
- Clonidine – sexual dysfunction
- Methyldopa – depression
- Thiazide diuretics – hyperglycemia
- Fibrates – Kidney injury
- Statins – liver damage

Consequences of The Illusion of Control

Introduction

Uncertainty

The Illusion
of Control

Risk

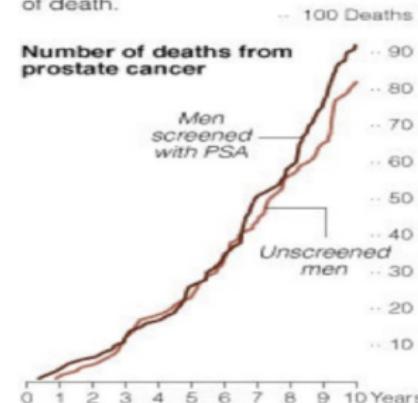
Exercise

PSA TESTS FOR PROSTATE CANCER

New England Journal of Medicine, March 2009

Overdiagnosis and overtreatment!

A study of 77,000 American men found that those who received annual PSA blood tests for prostate cancer did not have a reduced rate of death.



Bad Luck

Boris
Babic,
HKU

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

Long life's due to genes, says expert

LONDON: People who want to live to 100 should not worry about their lifestyle because whether you do or do not is likely to be determined by your genes, according to a longevity expert.

By and large, centenarians can live "unhealthy" lifestyles and get away with it, said Professor Nir Barzilai, director of the Institute of Ageing at the Albert Einstein College of Medicine in New York.

Those who live very long lives are genetically programmed to do so; it insulates them from the effects of "environmental" factors like smoking and a poor diet, he was quoted as saying by Britain's Telegraph newspaper yesterday.

Living a healthy life might help increase life expectancy by a few years, but it would not help those who want to live much longer, he said.

The Telegraph said the professor studied 500 Jewish people between 95 and 112 years of age.

He was scheduled to address the Royal Society in London yesterday on the subject of "interfering with ageing to treat ageing-related disease".

Speaking ahead of his lecture, Prof Barzilai told the newspaper that centenarians tended to have genes which delayed the onset of age-related illnesses like heart disease and Alzheimer's.

"When they eventually die, they die of the same things that people die of in their 70s or 80s," he said. "It's just that they die 30 years later."

Identifying these genes opened the doorway to developing longevity drugs which mimicked their effects, he said. Prof Barzilai and his team have already identified a number of such genes.

Laboratories are now working to create a drug which mimics the effects of three of them - two that increase the production of so-called "good" cholesterol in the body, which reduces the risk of heart disease and stroke, and a third that helps prevent diabetes.

Bad Luck

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

PREDICTIONS: MEDICINE

Cancer: Lifestyle and Genetics or Bad Luck?

While it sometimes does come down to lifestyle and genetics, researchers have found that most of the time it is something else entirely: a simple matter of (bad) luck. According to a mathematical analysis by researchers at Johns Hopkins School of Medicine, 65% of the cancer risk is due to random mistakes in a cell's DNA that can occur every time it divides (i.e., bad luck).

Science, Jan 2, 2015

How to Manage Illusion of Control

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

PARADOX OF CONTROL: THE ILLS OF PILLS

- **The Point is not that medicine is not useful**
- **The Point is avoiding the Illusion of Control, that is accepting the Paradox of Control leading to more control on health and longevity, i.e.,**
- **Favor Evidence-Based Medicine.**

Much of the gain in life expectancy over the last 100 years or so can be attributed to better sanitation, hygiene, education, quarantine, etc. For example,

- In US cities, between 1900 and 1936, 43% decline in mortality resulted from water filtration and chlorination
(David Pilling, FT: November 7 2018)

How to Manage Illusion of Control

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

ILLUSION OF CONTROL

Underestimating uncertainty, or worse, believing that we can influence chance events by our actions



Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

Risk

Subjective vs. Objective Risks

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

- We will often encounter a distinction between subjective and objective uncertainty.
- Sometimes the word risk is used for objective uncertainty, and the word uncertainty is reserved for subjective uncertainty.
- Objective uncertainty is quantifiable. Subjective uncertainty is not (??).
- But be careful: many things that may not seem quantifiable, are more quantifiable than you think. This is what the Bayesian approach will teach us.
- Anil Gaba: Subway uncertainty vs. Coconut Uncertainty.

Subjective vs. Objective Risks

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

THE STORY OF KLAUS:

Two types of Uncertainty:

“Subway Uncertainty” - can be modeled and quantified

“Coconut Uncertainty” - difficult if not impossible to model and quantify

Most real-life situations have a mixture of

Subway and Coconut Uncertainty

Subjective vs. Objective Risks

Introduction

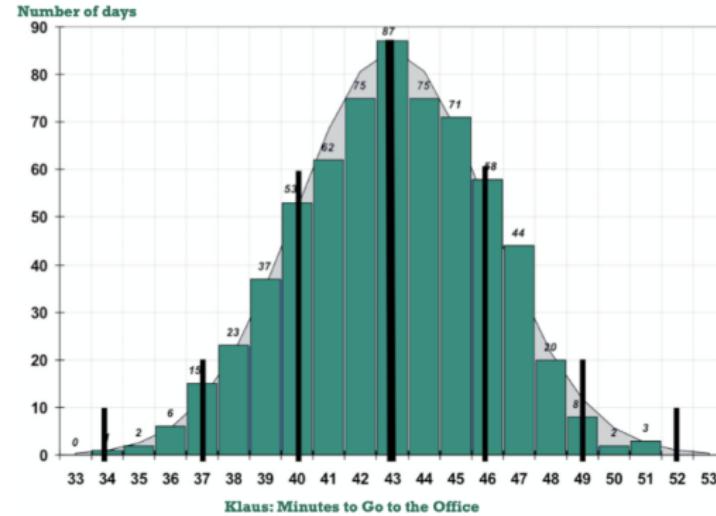
Uncertainty

The Illusion
of Control

Risk

Exercise

SUBWAY UNCERTAINTY



Introduction

Uncertainty

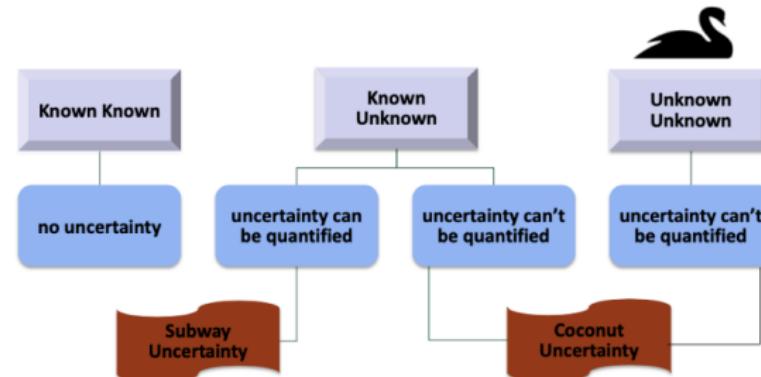
The Illusion
of Control

Risk

Exercise

Subjective vs. Objective Risks

SUBWAY + COCONUT UNCERTAINTY



Subjective vs. Objective Risks

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

FOCUS ON A GOOD JUDGMENT PROCESS



Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

Exercise

Exercise

Boris
Babic,
HKU

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

- We have approximately 100 students in class. For simplicity, assume it is exactly 100.
- What is the probability that at least one pair of students share a birthday?

Algebraic Solution

Boris
Babic,
HKU

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

$$\frac{100 \times 99}{2} = 4950 \text{ combinations of students}$$

$\frac{364}{365}$ probability that any single pair does not share a birthday

$\left(\frac{364}{365}\right)^{4950} \approx 0$ probability that no pair shares a birthday

$1 - (\approx)0 \approx 1$ probability that at least one pair does share a birthday

Finally Some Data Analysis

Introduction

Uncertainty

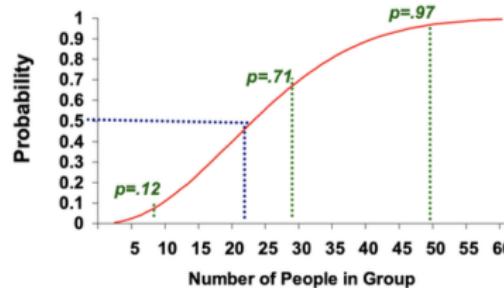
The Illusion
of Control

Risk

Exercise

```
p <- numeric(k) # create numeric vector to store probabilities
for (i in 1:k) {
    q <- 1 - (0:(i - 1))/365 # 1 - prob(no matches)
    p[i] <- 1 - prod(q) }
prob <- p[k]
print(prob)
```

PROBABILITY OF AT LEAST ONE MATCH IN GROUP OF SIZE N



The R Statistical Programming Language

Introduction

Uncertainty

The Illusion
of Control

Risk

Exercise

- For computing purposes, we will use the statistical programming language **R**.
- **R** is an open source language, written in C and Fortran.
- It was initially developed at Bell Labs, where it was known (creatively) as S.
- Today it is the most widely used language among statisticians.
- You can download a free distribution on the **R** project web page (www.r-project.org/) together with RStudio (rstudio.com) which is the leading IDE for **R**.